

Conference Abstract

Data Infrastructure for Scientific and Collections Data in Medium Size Institution

Jiri Frank[‡], Jakub Belka[‡][‡] National Museum, Prague, Czech RepublicCorresponding author: Jiri Frank (jiri_frank@nm.cz)

Received: 18 Aug 2017 | Published: 18 Aug 2017

Citation: Frank J, Belka J (2017) Data Infrastructure for Scientific and Collections Data in Medium Size Institution. Proceedings of TDWG 1: e20429. <https://doi.org/10.3897/tdwgproceedings.1.20429>

Abstract

We in the National Museum in Prague were recently faced with a difficult decision. The question was how to build up optimal and economical infrastructure to store and back up our scientific and collections data, which currently requires approximately 150 TB of data storage and is growing constantly. In addition to the infrastructure it was also important to consider a potential LTP (Long Term Preservation) solution. We would like to share our experience and maybe inspire you with our model. The infrastructure model can be defined by four main elements: visualisation of the data infrastructure, virtual platform, back-up and LTP. The visualisation is done by constantly updating a data schema that shows the data stores and their connections with virtual platforms. Every data store has a defined data structure. For example data storage for collections data reflects their physical structure, location and distribution. So it creates a virtual collections depository divided in collections and sub-collections on various levels. For our virtualisation platform we chose the solution by VMware. This platform creates a data space from high speed local data stores. This space is used for various database systems in the museum, e.g. for collections management. Those database systems are connected with large capacity lower speed data stores. The infrastructure is designed to guarantee fast access to the databases and metadata with lower requirements for a storage capacity. The access to the digitised master files (images etc.) is indeed a bit slower due to the lower speed large capacity storage volumes. The back-up strategy has two options. We are using for the virtualisation platform and virtual machines VEEAM back-up system, which works on the basis of

reverse incremental backup. The images of virtual machines are backed up on external data storage in a data centre (hosted by third party). The back-up of large capacity lower speed data stores is done by incremental back up to directly connected external data stores. The external data storage in the data centre is replicated in two separate geographical locations. For the future we are planning an LTP strategy for data and also metadata. The best technology at the moment with high capacity, reasonable price, and a long preservation (more than 100 years) is an Optical Disc Archive (ODA). One of the advantages of this technology is the lack of special requirements for temperature, humidity, etc., as well as economical space requirements. The whole system with the LTP solution and technical descriptions will be described as a schema on the poster.

Keywords

Data visualisation, virtualisation, back up, Long Term Preservation

Presenting author

Jakub Belka